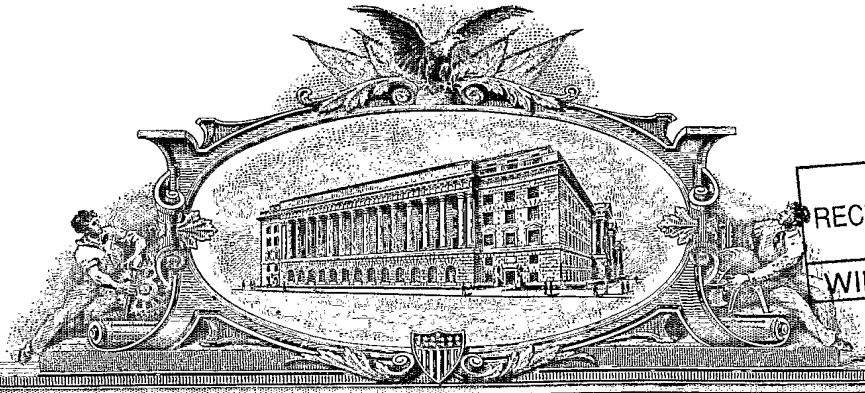

P1 1183232

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

June 17, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/470,544*
FILING DATE: *May 09, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/07733*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

N. WOODSON
Certifying Officer

Express Mail Label No.: EV139502880US
Date of Deposit: May 9, 2003

Attorney Docket No. 21402-646 (CURA-946)

ionPROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR § 1.53(c)

INVENTORS / APPLICANTS			
Given Name (last name, first name, and middle initial [if any])			Residence (City and State or Foreign Country)
Kekuda	Ramesh		Norwalk, CT 06851

TITLE OF THE INVENTION

NOVEL NOVEL HUMANIN VARIANT WITH L12S, A24T AND S14T, E15A AND 116T REPLACEMENTS-LIKE PROTEINS AND NUCLEIC ACIDS ENCODING SAME

CORRESPONDENCE ADDRESS

Attorney Name: Firm Name and Address:	Ivor R. Elrifi, Ph.D. MINTZ, LEVIN, COHN, FERRIS, GLOVSKY AND POPEO, P.C. One Financial Center Boston, MA 02111
Telephone:	(617) 542-6000
Fax:	(617) 542-2241

ENCLOSED APPLICATION PARTS

☒ **Specification**	Number of Pages: 10
☐ **Sequence Listing**	Number of Pages:
☐ **Drawings** (☐ Formal; ☐ Informal)	Number of Sheets:

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government:

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are:

METHOD OF PAYMENT

☒ A check in the amount of **$160.00** is enclosed to cover the filing fees of the Provisional application.

☐ The Commissioner is hereby authorized to charge **$160.00** is enclosed to cover the filing fees of the Provisional application to Deposit Account No. 50-0311, Reference No. _________.

☒ The Commissioner is hereby authorized to charge additional fees or credit any overpayment to Deposit Account No. 50-0311, Reference No. 21402-646 (CURA-946)

Respectfully submitted,

Reg. No. 44,848

Date: May 9, 2003

Ivor R. Elrifi, Registration No. 39,529
MINTZ, LEVIN, COHN, FERRIS, GLOVSKY AND POPEO, P.C.
One Financial Center
Boston, MA 02111
Telephone: (617) 542-6000

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT
SEND TO: ***BOX PROVISIONAL APPLICATION, Assistant Commissioner for Patents, Washington, DC 20231***

Please address all written correpondence to customer number 30623.

TRA 1793659v1

30623
PATENT TRADEMARK OFFICE

Novel Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like Proteins and Nucleic Acids Encoding Same

The present invention discloses a novel protein encoded by a cDNA and/or by genomic DNA and proteins similar to it, namely, new proteins bearing sequence similarity to Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements, nucleic acids that encode these proteins or fragments thereof, and antibodies that bind immunospecifically to a protein of the invention.

Background

"OMIM Number - 606120

Hashimoto et al. (2001) noted that important clues in the development of therapy for Alzheimer disease (FAD; 104300) come from the study of molecules that suppress FAD gene-induced death in neuronal cells in culture. Using the death-trap screening method devised by Vito et al. (1996), they identified a cDNA, which they called humanin (HN), encoding a deduced 24-amino acid secretory polypeptide that suppresses neuronal cell death induced by 3 FAD genes: amyloid precursor protein (APP; 104760), presenilin-1 (PS1; 104311), and presenilin-2 (PS2; 600759). The peptide also abolished death caused by A-beta amyloid, but had no effect on death by Q79 or superoxide dismutase-1 mutants. Transfected HN cDNA was transcribed to the corresponding polypeptide and then was secreted into the cultured medium. The rescue action clearly depended on the primary structure of HN. Northern blot analysis detected expression of major 1.6- and minor 3.0- and 1.0-kb transcripts at high levels in heart, skeletal muscles, kidney, and liver, at lower but significant levels in brain and the gastrointestinal tract, and at barely detectable levels in the immune system.

The cDNA sequence of HN (GenBank AY029066), is 99% identical to the sequence of 16S mitochondrial ribosomal RNA (561010), which is mitochondrially encoded, but is also 99% identical to some nuclear-encoded cDNAs. It was therefore not clear whether the HN cDNA is mitochondrial ribosomal RNA or a nuclear-transcribed mRNA. "

An additional locus was found in human genomic sequence which encode humanin like polypeptides on chromosome 6 (See the table below).

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END
browser details	YourSeq	13	1	23	24	78.3%	6	++	62235629	62235697

Locus on chromosome 6 encodes another humanin related polypeptide (CG202524-08) which has 5 amino acid replacements - S12L and T24A, S14T, E15A and I16T as compared to the known humanin GenBank AY029066.

```
CG202524-02  1  MAPRGFSCLLLSTSEIDLPVKRRA----  24
AY029066.1   1  MAPRGFSCLLLLTSEIDLPVKRRA----  24
CG202524-04  1  MAPRGFSCLLLSTSEIDLPVKRLLSSVF  28
CG202524-03  1  MAPRGFSCLLLSTSEIDLPVKRRT----  24
CG202524-08  1  MARRGFSCLLLSTTATDLPVKRRT----  24
```

An interesting observation made when the the novel CuraGen polypeptide was compared to the known Humanin (AY029066.1) as well as three other novel CuraGen pepties (CG202524-02, CG202524-03 and CG202524-04) is that Leu at position 12 was replaced by Ser. Hashimoto et al (2001) (reference 4) have done a systematic site-directed mutagenesis analysis of Humanin and have identified that replacing Leu-12 with Ala abolished the protective function of Humanin. The observation that there are 5 genomic loci which encode a Ser in place of Leu at position 12 might indicate a functional significance to this amino acid. For example, Hashimoto et al (2001) (reference 4) have shown that Ser14 substitution to Gly caused potentiated the neuroprotective activity of Humanin thousand fold, whereas the substitution to Ala nullified the protective activity. It is conceivable that substitution of Leu12 with Ser (found in 5 different genetic loci and found in the 3 novel Humanin variants described here) potentiates the neuroprotective activity of humanin. It is also possible that the replacement of Ala 24 with Thr in –08 variant and S14T, E15A and I16T in the –08 variant might have beneficial potentiating neuroprotective activities.

REFERENCES

1. Hashimoto, Y.; Niikura, T.; Tajima, H.; Yasukawa, T.; Sudo, H.; Ito, Y.; Kita, Y.; Kawasumi, M.; Kouyama, K.; Doyu, M.; Sobue, G.; Koide, T.; Tsuji, S.; Lang, J.; Kurokawa, K.; Nishimoto, I. :
A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and A-beta. Proc. Nat. Acad. Sci. 98: 6336-6341, 2001.
PubMed ID : 11371646

2. Vito, P.; Lacana, E.; D'Adamio, L. :
Interfering with apoptosis: Ca(2+)-binding protein ALG-2 and Alzheimer's disease gene ALG-3. Science 271: 521-524, 1996.
PubMed ID : 8560270

3. Hashimoto Y, Niikura T, Ito Y, Sudo H, Hata M, Arakawa E, Abe Y, Kita Y, Nishimoto I. :
Detailed characterization of neuroprotection by a rescue factor humanin against various Alzheimer's disease-relevant insults.
J Neurosci. 2001 Dec 1;21(23):9235-45

Brief Description of the Drawings

Figure 1. Nucleotide sequence encoding the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein (Acc. No. CG202524-08) of the invention.

Figure 2. Protein sequence encoded by the nucleotide sequence shown in Figure 1.

Figure 3A. A high-scoring match as determined by a BLASTN search of GenBank Composite (no HTG) dated 05/08/03 using the sequence of the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like gene of the invention.

Figure 3B. A high-scoring match as determined by a BLASTP search (versus Non-Redundant Composite dated 05/08/03) using the sequence of the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein of the invention.

Figure 3C. BLASTN identity search of CuraGen Corporation's human SeqCalling database using the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like gene of the invention.

Figure 4. ClustalW alignment of the protein of Acc. No. CG202524-08 with similar Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacementss.

Figure 5: PSORT, SignalP and hydropathy results for the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein of Acc. No. CG202524-08.

Description of the Invention

Method of Identifying the Nucleic Acid Encoding the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-Like Protein.

The sequence of Acc. No. CG202524-08 was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. *In silico* prediction was based on sequences available in Curagen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression,

physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

One or more genomic clones AL356135.11 on chromosome 6 were identified by TBLASTN using CuraGen Corporation's sequence file for members of Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements and/or the Humanins family, run against the genomic daily files made available by GenBank or obtained from Human Genome Project Sequencing centers. These sequences were analyzed for putative coding regions as well as for similarity to known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs. Putative coding regions were spliced from the genomic clone and then concatenated using a known homolog for reference. The derived sequence may have been further extended using additional genomic clones showing greater than 98% identity to the open reading frame.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze genomic clones was reiterated to derive the full length sequence. The following public components were thus included in the invention: AL356135.11.

The DNA sequence was analyzed to identify any open reading frames encoding novel full length proteins as well as novel splice forms of these genes. The DNA sequence and protein sequence for a novel Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like gene are reported here as CuraGen Acc. No. CG202524-08.

Results

The novel nucleic acid of 75 nucleotides (designated CuraGen Acc. No. CG202524-08) encoding a novel Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein is shown in Fig. 1. An open reading frame was identified beginning at nucleotides 1-3 and ending at nucleotides 73-75. This open reading from begins with an ATG initiation codon and ends with a TAA stop codon. This polypeptide represents a novel functional Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein. The start and stop codons of the open reading frame are highlighted in bold type. Putative untranslated regions (underlined), if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 24 amino acid residues is presented using the one-letter code in Fig. 2.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 70 of 75 bases (93%) identical to a gb:GENBANK-ID:AF227907|acc:AF227907.1 mRNA from Homo sapiens (Homo sapiens chromosome 17 sequence containing mitochondrial genome insertion) (Fig. 3A). The full amino acid sequence of the protein of the invention was found to have 18 of 23 amino acid residues (78%) identical to,

and 19 of 23 amino acid residues (82%) similar to, the 24 amino acid residue ptnr:SPTREMBL-ACC:Q8IVG9 protein from Homo sapiens (Human) (Humanin)(Fig. 3B).

A multiple sequence alignment is given in Fig. 4, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized in Table 1.

No significant domains found

This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like gene disclosed in this invention maps to chromosome 6. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue expression

The Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like gene disclosed in this invention is expressed in at least the following tissues: not available. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG202524-08. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF227907|acc:AF227907.1) a closely related Homo sapiens chromosome 17 sequence containing mitochondrial genome insertion homolog in species Homo sapiens : heart, skeletal muscles, kidney, and liver, at lower but significant levels in brain and the gastrointestinal tract .

Cellular Localization and Sorting

The PSORT, SignalP and hydropathy profile for the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein are shown in Fig. 5. Although PSORT suggests that the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein may be localized extracellularly, the protein of CuraGen Acc. No. CG202524-08 predicted here is similar to the Humanins family, some members of which are secreted. Therefore it is likely that this novel Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein is localized to the same sub-cellular compartment.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein includes the nucleic acid whose sequence is provided in Fig. 1, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Fig. 1 while still encoding a protein that maintains its Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of CuraGen Acc. No. CG202524-08, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 7% of the bases may be so changed.

The novel protein of the invention includes the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein whose sequence is provided in Fig. 2. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Fig. 2 while still encoding a protein that maintains its Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 22% of the amino acid residues may be so changed.

Chimeric and Fusion Proteins

The present invention includes chimeric or fusion proteins of the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein, in which the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein of the present invention is joined to a second polypeptide or protein that is not substantially homologous to the present novel protein. The second polypeptide can be fused to either the amino-terminus or carboxyl-terminus of the present CG202524-08 polypeptide. In certain embodiments a third nonhomologous polypeptide or protein may also be fused to the novel Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein such that the second nonhomologous polypeptide or protein is joined at the amino terminus, and the third nonhomologous polypeptide or protein is joined at the carboxyl terminus, of the CG202524-08 polypeptide. Examples of nonhomologous sequences that may be incorporated as either a second or third polypeptide or protein include glutathione S-transferase, a heterologous signal sequence fused at the amino terminus of the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein, an immunoglobulin sequence or domain, a serum protein or domain thereof (such as a serum albumin), an antigenic epitope, and a specificity motif such as $(His)_6$.

The invention further includes nucleic acids encoding any of the chimeric or fusion proteins described in the preceding paragraph.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, $(Fab)_2$ or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein may have important structural and/or physiological functions characteristic of the Humanins family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Cardiomyopathy, Atherosclerosis,Hypertension, Congenital heart defects, Aortic stenosis ,Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus , Pulmonary stenosis , Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity,Transplantation, Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalceimia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Cirrhosis as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

FIGURES

Figure 1. Nucleotide sequence encoding the Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein of the invention.

```
>CG202524-08 Novel humanin variant with L12S,  A24T and S14T, E15A and I16T replacements,
submitted to study DDPT on 05/08/03 by rkekuda; classification type=Finished In-silico;
novelty=Novel; ORF start=1, ORF stop=73, frame=1; 75 bp, 75 bp (DNA).
ATGGCTCGACGAGGTTTCAGCTGTCTCTTACTTTCAACCACTGCAACTGACCTGCCCGTG   60
AAGAGGCGGACATAA   75
```

Figure 2. Protein sequence encoded by the nucleotide sequence shown in Figure 1.

```
>CG202524-08
MARRGFSCLLLSTTATDLPVKRRT   24
```

Figure 3A. BLASTN search using CuraGen Acc. No. CG202524-08.

```
>gb:GENBANK-ID:AF227907|acc:AF227907.1 Homo sapiens chromosome 17 sequence
           containing mitochondrial genome insertion - Homo sapiens, 14722 bp.
           Length = 14,722

 Plus Strand HSPs:

 Score = 330 (49.5 bits), Expect = 1.3e-07, P = 1.3e-07
 Identities = 70/75 (93%), Positives = 70/75 (93%), Strand = Plus / Plus

Query:     1 ATGGCTCGACGAGGTTTCAGCTGTCTCTTACTTTCAACCACTGCAACTGACCTGCCCGTG 60
             ||||||| |||||| ||||||||||||||||||||||||| || || ||||||||||||
Sbjct:  5353 ATGGCTCCACGAGGGTTCAGCTGTCTCTTACTTTCAACCAGTGAAATTGACCTGCCCGTG 5412

Query:    61 AAGAGGCGGACATAA 75
             |||||||||||||||
Sbjct:  5413 AAGAGGCGGACATAA 5427
```

Figure 3B. BLASTP search using the protein of CuraGen Acc. No. CG202524-08.

```
>ptnr:SPTREMBL-ACC:Q8IVG9 Humanin - Homo sapiens (Human), 24 aa.
           Length = 24

 Score = 87 (30.6 bits), Expect = 0.0014, P = 0.0014
 Identities = 18/23 (78%), Positives = 19/23 (82%)

Query:     1 MARRGFSCLLLSTTATDLPVKRR 23
             MA RGFSCLLL T+  DLPVKRR
Sbjct:     1 MAPRGFSCLLLLTSEIDLPVKRR 23
```

Figure 3C. BLASTN identity search of CuraGen Corporation's Human SeqCalling database using CuraGen Acc. No. CG202524-08.

```
>s3aq:248012199 , 580 bp.
           Length = 580

 Minus Strand HSPs:

 Score = 312 (46.8 bits), Expect = 2.0e-08, P = 2.0e-08
 Identities = 68/75 (90%), Positives = 68/75 (90%), Strand = Minus / Plus

Query:    75 TTATGTCCGCCTCTTCACGGGCAGGTCAGTTGCAGTGGTTGAAAGTAAGAGACAGCTGAA 16
             ||||| ||||||||||||||||||||||| || || ||||| ||||||||||||||||||
Sbjct:   257 TTATGCCCGCCTCTTCACGGGCAGGTCAATTTCACTGGTTAAAAGTAAGAGACAGCTGAA 316

Query:    15 ACCTCGTCGAGCCAT 1
             ||||||| |||||||
Sbjct:   317 CCCTCGTGGAGCCAT 331
```

Figure 4. ClustalW alignment of CG202524-08 protein with related proteins.

```
CG202524-02  1  MAPRGFSCLLLSTSEIDLPVKRRA----  24
AY029066.1   1  MAPRGFSCLLLLTSEIDLPVKRRA----  24
CG202524-04  1  MAPRGFSCLLLSTSEIDLPVKRLLSSVF  28
CG202524-03  1  MAPRGFSCLLLSTSEIDLPVKRRE----  24
CG202524-08  1  MARRGFSCLLLSTIATDLPVKRRT----  24
```

Information for the ClustalW proteins:

Accno	Common Name	Length
CG202524-08	novel Novel humanin variant with L12S, A24T and S14T, E15A and I16T replacements-like protein	24
CG202524-02	Humanin L12S variant	24
AY029066.1	Homo sapiens Humanin (HN1) mRNA, complete cds.	24
CG202524-03	Novel humanin variant with L12S and A24T replacements	24
CG202524-04	Humanin like gene with L12 S, R23L, and A24L replacements and SSVF insertion at aa 25 to 28	28

In the alignment shown above, black outlined amino acid residues indicate residues identically conserved between sequences (i.e., residues that may be required to preserve structural or functional properties); amino acid residues with a gray background are similar to one another between sequences, possessing comparable physical and/or chemical properties without altering protein structure or function (e.g. the group L,V, I, and M may be considered similar); and amino acid residues with a white background are neither conserved nor similar between sequences.

Figure 5: PSORT, SignalP and hydropathy results for CuraGen Acc. No. CG202524-08.

```
mitochondrial intermembrane space --- Certainty=0.8420(Affirmative) < succ>
       mitochondrial matrix space --- Certainty=0.6797(Affirmative) < succ>
     mitochondrial inner membrane --- Certainty=0.3682(Affirmative) < succ>
     mitochondrial outer membrane --- Certainty=0.3682(Affirmative) < succ>


Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    24        0.074   0.37   NO
  max. Y    18        0.197   0.34   NO
  max. S     1        0.907   0.88   YES
  mean S     1-17     0.679   0.48   YES
# Most likely cleavage site between pos. 17 and 18: ATD-LP
```